



1970

An Investigation of the Relationship between Art Quality and Clinical Judgment on the Draw-A-Person Test

James Clarence Young
Loyola University Chicago

Recommended Citation

Young, James Clarence, "An Investigation of the Relationship between Art Quality and Clinical Judgment on the Draw-A-Person Test" (1970). *Dissertations*. Paper 1238.
http://ecommons.luc.edu/luc_diss/1238

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License](https://creativecommons.org/licenses/by-nc-nd/3.0/).
Copyright © 1970 James Clarence Young

An Investigation of the Relationship Between
Art Quality and Clinical Judgment
on the Draw-A-Person Test

by
James C. Young

A Dissertation Submitted to the Faculty of the
Graduate School of Loyola University in
Partial Fulfillment of the Require-
ments for the Degree of
Doctor of Philosophy

September

1970

Life

James Clarence Young was born in Chicago, Illinois on January 26, 1943.

He graduated from Quigley Preparatory Seminary, Chicago, in June, 1961. In June, 1963, he graduated from St. Mary of the Lake Junior College in Niles, Illinois. Majoring in psychology, he received the degree of Bachelor of Science from Loyola University, Chicago, in June, 1965.

He was awarded an A. J. Schmitt Fellowship and began full-time graduate study in the Department of Psychology at Loyola University in September, 1965. He took his clerkship at Hines Veterans Administration Hospital. His internship was started at Hines and completed at the Veterans Administration West Side Hospital, Mental Hygiene Clinic. From September of 1968 until June of 1970, he served as a clinical assistant to the Director of Clinical Training at Loyola University.

Acknowledgments

The author wishes to express his gratitude to Doctor Frank J. Kobler for the advice, interest, encouragement, and support he lent to this research. Many thanks are extended also to Doctors Ronald E. Walker and Roderick W. Pugh who offered valuable suggestions on methodology and treatment of the data. Finally, the author extends a special thanks to all those who participated in this research: to the subjects, both college students and patients, without whose cooperation this study could not have been completed; to the art judges who took time out from their busy schedules to evaluate the drawings; and to those graduate students who, in spite of the many demands on them occasioned by their own research projects, freely gave of their time to serve as clinical judges.

Table of Contents

Chapter	Page
I. Introduction	1
II. Review of Related Literature	6
III. Procedure.	26
IV. Results.	38
V. Discussion	63
VI. Summary.	74
Abstract	78
References	79

List of Tables

Tables	Page
1. Intraclass Reliabilities of Ratings for Clinical and Artistic Judge Groups.	39
2. Analysis of Variance of Adjustment Ratings for Naive Clinical Judges.	41
3. Analysis of Variance of Adjustment Ratings for Informed Clinical Judges	42
4. Analysis of Variance of Adjustment Ratings for Combined Clinical Judge Groups	43
5. Proportion of Correct Patient-Nonpatient Judgments for Two-Drawing and Five-Drawing Series. .	48
6. Correlations of Judges' Degree of Confidence with Assignment of Patient-Nonpatient Status and Correctness of Judgment.	49
7. F Values and Correlation Ratios of Judges' Confidence Ratings for Drawings from Four Levels of Artistic Quality	50
8. Intraclass Reliability of Adjustment Ratings for Three Levels of Artistic Quality: Poor Art, Average Art, and Good Art.	55
9. Comparison of Intraclass Reliabilities of Adjustment Ratings for Three Levels of Artistic Quality: Poor Art (P.A.), Average Art (A.A.), and Good Art (G.A.).	56
10. Significance Tests for the Proportion of Correct Patient-Nonpatient Judgments for Three Levels of Artistic Quality: Poor Art, Average Art, and Good Art	57
11. Comparison of Proportion of Correct Patient-Nonpatient Judgments for Three Levels of Artistic Quality: Poor Art (P.A.), Average Art (A.A.), and Good Art (G.A.).	58

Chapter I

Introduction

It has long been a common sense observation that a work of art often reveals a good deal about the personality of the one who produced it. Novelists, for example, have frequently been seen as depicting themselves in the characters of their heroes; poets as proclaiming their own desires and passions in a near monothematic series of poems; and painters as revealing the nature and intensity of their emotional states through the selection of color or the quality of a brush stroke. With some, such as Van Gogh, the presence of personality factors was clear and unmistakable. For others, such as Shakespeare, the expression of such characteristics was definitely more subtle resulting, not infrequently, in conflicting reports by those who have attempted to analyze the author.

Thus the notion of personality projection in drawings is certainly not a new idea. The contributions of Machover (1949) and Buck (1948) lay precisely in the fact that they developed drawing techniques whose primary (and only) purpose was one relating to the evaluation of intellectual and personality functioning. In addition to standardizing techniques and formulating sets of specific interpretative rules, they popularized the notion of the body image, a concept referring to the individual's perception of his own body and its relation to the social and physical environment in which he lives. More specifically, Hammer (1968) has summarized the theoretical underpinnings of

projective drawings in the following three postulates:

- (a) There is a tendency in man to view the world in an anthropomorphic manner; that is, in his own image.
- (b) The core of the antropomorphic view of the environment is the mechanism of projection.
- (c) Distortions enter into the process of projection to the extent to which the projection has a defensive function; that is, the projection is in the service of ascribing to the outer world that which the subject denies in himself (p. 369).

Clinicians who employed the projective drawings were favorably impressed with their apparent utility, and before long the Draw-A-Person Test (DAP) became a routinely administered part of most psychological test batteries. Within time, other techniques were developed to supplement the original instrument. Though most still relied strongly on the body-image hypothesis, many aimed in addition at the detection of psychologically relevant factors within more or less circumscribed areas. Caligor's (1951, 1952, 1953) Eight-Card-Redrawing Test, for example, boasted of having higher validity and reliability than the parent instrument and proclaimed an especial competence in the area of psychosexual identification. Similar in their specialized intent were Harrower's (1950) Most Unpleasant Concept Test, Kinget's (1952) Drawing Completion Test, and Hammer's (1958) Chromatic House-Tree-Person Test. Finally, there were those techniques of uncertain origin which probably grew out of clinical experience and were passed along by word of mouth. Among the better known were the draw-a-family and draw-yourself techniques, the Draw-An-Animal Test (which professed to tap the biological as opposed to the social side of man), and the Draw-A-Person-in-the-Rain variation (which

sought a picture of defensive patterns under conditions of environmental stress). Unfortunately, none of these techniques have received the research attention they deserve. Consequently, an adequate estimation of their respective assets and liabilities is impossible to obtain at the present time.

In contrast to the paucity of research surrounding the use of the supplementary drawing techniques, the DAP itself has been the object of numerous investigations. Nonetheless, the user of the DAP seeking support for his beliefs has been able to derive little satisfaction from a perusal of the literature. For one thing, much of the earlier research was conducted with little or no regard for even the most elementary methodological principles. To make matters worse, the conclusions derived from the better designed studies were often diametrically opposed; and it was rare when research results lent support to one of Machover's hypotheses in a consistent and unequivocal manner.

Following Swenson's (1957) thorough and largely unfavorable review of the status of the DAP, it seemed likely that both research interest and clinical usage of the test would be certain to decline. Neither seems to have occurred. Perhaps the persistent popularity of the test derives from the ease with which it can be routinely administered and interpreted. Or perhaps, as Swenson (1957) himself suggested, it may be that clinicians have been the unwilling victims of a partial reinforcement schedule whereby the dynamic insights afforded by an occasionally spectacular DAP are encoded in the user's mind long after the dis-

appointments of the more numerous failures of the test have been forgotten. In any case, Sundberg (1961) has reported that the DAP ranks second only to the Rorschach in terms of frequency of usage among practicing clinicians. As for research, not only has a wide variety of studies continued to be published (Roback, 1968; Swenson, 1968), but recent investigations have evinced a type of procedural and statistical sophistication generally found wanting throughout the earlier literature. Though definitive statements concerning most DAP hypotheses must await further research, sufficiently consistent evidence has been accumulating in some areas so that certain general conclusions can now be seen as emerging.

One such area of remarkably consistent findings is represented by those studies which have shown that users of the test, in making their clinical evaluations, appear to be more strongly influenced by the artistic quality of a DAP protocol than by the presence or absence of any of the traditional pathognomic indicators (Roback, 1968; Swenson, 1957, 1968). Typically, 35% to 65% of the variance associated with clinical judgments of adjustment has been accounted for by an art-quality factor alone. It seems difficult to overestimate the importance of these findings. It is somewhat puzzling, therefore, that since the influence of artistic quality was first discovered (Whitmyre, 1953), only a handful of studies have bothered taking this factor into account. Further, literally no studies to date have attempted to see what techniques, if any, could be found which might serve to attenuate a clinician's tendency to respond primarily to the

art quality of a drawing. This last consideration seems most important; for it seems likely that the DAP will eventually become of historical interest only unless such techniques can be devised. It was thus the purpose of this study to investigate more fully the relationship between artistic quality and clinical judgment and to attempt to determine if there were any means by which the influence of art quality could be lessened.

The major hypotheses of this study may be stated as follows:

Hypothesis I: that clinical judgments of pathology are the result of an interaction between S's level of adjustment and the artistic quality of his drawing ($p = .05$).

Hypothesis II: that pathology results in poor art to the extent the object drawn is personally meaningful to the individual ($p = .05$).

Hypothesis III: that providing clinicians with research findings concerning the DAP will increase the accuracy of their judgments ($p = .05$).

Hypothesis IV: that providing the clinician with an extended series of drawings will enable him to increase the accuracy of his judgments ($p = .05$).

Hypothesis V: that there is a positive correlation between a clinician's confidence in his judgment and the correctness of that judgment ($p = .05$).

Chapter II

Review of Related Literature

Perhaps the concept most central to the DAP has been the body-image hypothesis, the belief that the figures drawn reflect in a meaningful way some important aspects of the personality of the S doing the drawings (Machover, 1949). In general, it may be said that the literature to date, while not conclusive, does offer a measure of support for this most basic notion. Wysocki and Whitney (1965), for example, found it possible to successfully differentiate crippled children from normals on the basis of the drawings done by each group. Similar results were earlier obtained with hemiplegic patients by Prater (1950), with obese women by Katkov and Goodman (1953), and by Schmidt and McGowan (1959) who employed physically handicapped Ss. The latter investigators further noted that such differentiations are done poorly by "cognitive" judges but performed quite successfully by judges more "affectively" oriented. In this study, the "cognitive" judge was one who relied on specific signs or factors in his analysis while the "affective" judges took an impressionistic, feeling approach to the DAP. Finally, and perhaps most convincing, was the recent study by Apfeldorf and Smith (1966) who obtained photographs of their Ss and found judges able to match the photos of the person with the drawings rendered by him. The authors note, however, that their findings, while statistically significant, would be only of limited value in decisions involving the individual case.

The difficulty in evaluating research on the body-image hypothesis is, as Swenson (1968) has noted, that the construct validity of the term has never been adequately established. What, in other words, is the criterion of the body image against which results might be compared? The above studies, for example, seem to imply that the best criterion is the person as he really looks. But is this the only measure? What of the person as he psychologically feels and senses himself to be? That is, might the body image be much the same as the self-concept? In line with this notion and adding to the confusion has been Hammer's (1959, 1968) contention that not only the "real" self, but the ideal self, the feared self, the social self, the childhood self, and so on may each be represented in the drawing. Lending support to this latter notion was the study by Kamano (1960) which found that while 36% of the variance of drawing performances could be attributed to the actual self-concept, the ideal and least liked self-concepts accounted for 12% and 13% respectively. Again, however, one is faced with how to evaluate in a practical fashion the relative contributions of each of these measures in the drawings of a single individual. To date, such diverse criteria of body imagery as the GSR (Fisher, 1959), photographs (Apfeldorf & Smith, 1966), height and weight (Silverstein & Robinson, 1961), and the Secord-Jourard Body Cathexis Scale (Hunt & Feldman, 1960) all have been employed at one time or another. Thus, though it seems that the body-image hypothesis contains some measure of truth, it is obvious that any final decisions concerning it must

await a clearer definition of the underlying construct involved.

A less confused pattern of research findings that has begun to emerge is in regard to the significance of the sex of the first drawn figure. Traditionally, it was Machover's (1949) contention that sexually inverted individuals might be distinguished by their tendency to draw as their first figure a member of the opposite sex. By as early as 1957, however, it had become obvious that this hypothesis, while containing an element of truth, was not one which was practically helpful to the clinician (Brown & Tolor, 1957; Swenson, 1957). Reasons for this lack of utility have become clearer with more recent research. Firstly, it was found that while homosexuals did draw opposite sex figures first to a greater extent than did normals, the difference was of no practical utility in evaluating the individual patient. Vilhoute (1955), for example, found that while this tendency was present in 24% of a sexually inverted group, it also characterized 18% of his normal controls. Secondly, it soon became clear that the sex of the first drawn figure was a rather complex phenomenon related, at least, to the age and sex of the S as well as to his state of adjustment (Butler & Marcuse, 1959; Craddick, 1963; Hammer & Kaplan, 1964; Swenson & Newton, 1955). To summarize these findings, it may be said that though boys and girls in the early school years tend to draw own-sex figures first, this tendency gradually increases with age for males but declines rather sharply for females (to as low

as 40% to 60% for college women). Though little cross-cultural research has been done in this area, a recent study by Millikan and Wahab (1969) suggests the phenomenon may be universal. Finally, and perhaps most significant, has been the finding that test-retest reliability studies have found a substantial amount of variability in the sex of the first drawn figure, especially for females (Litt & Margolies, 1966; Starr & Marcuse, 1959). While stability coefficients have typically been significant, it is not unusual for as many as 10% of the males and 40% of the females to change the sex of the first drawn figure. While this in itself may be a clinically relevant fact, research on this point is presently insufficient to permit the drawing of any conclusions.

In regard to the many other DAP hypotheses enumerated by Machover, research results have been almost uniformly non-supportive. In both the early (Jones & Thomas, 1961; Swenson, 1957) and more recent (Rovack, 1968; Swenson, 1968) reviews of the literature on the subject, the conclusion of "negative" or "inconsistent" findings was a theme reiterated again and again, especially in regard to Machover's content and structural hypotheses. Perhaps the most telling blow so far dealt to the DAP is to be found in the recently completed study of Wanderer (1969). In this study, drawings were collected from each of four diagnostic categories and a group of normals. Wanderer then presented sets of five drawings (one protocol from each group of Ss) to each of 20 DAP experts, including Karen Machover, Lauretta Bender, Helen

Anderson, Molly Harrower, Mildred Zadek, and Emanuel Hammer. The experts then assigned each of the five drawings to one of the five categories of Ss. Inspection of the data indicated that, while the experts correctly classified 40% of the drawings, this result was due to the near perfect classification of the drawings from the mentally defective group. When the data were re-analyzed omitting the judgments made of this group, the number of correct classifications fell to a level expected by chance alone. Finally, Wanderer failed to find a significant relationship between the reputed expertness of a judge (as determined by ratings given him by his peers) and the accuracy of his DAP judgments.

In defense of the clinical experts, Hammer (1969) has made several cogent criticisms of certain points of Wanderer's methodology. He notes, for example, that the failure to demonstrate a relation between expertness and clinical accuracy could be due either to the homogeneity of the judges' clinical skills or to the fact that the judges did not see the same set of five drawings. He further points out the rather arbitrary classification of parts of the data so that no differences appeared between the judges who got five out of five correct and those getting only two out of five correct. More serious still, the matching procedure employed in the study made it impossible for a judge to get only one out of five wrong; that is, if he were wrong on his first choice, he was automatically wrong on some other choice so that 40% wrong was the best he could do. Hammer concludes by noting that while only a minority of the experts considered the DAP their primary

diagnostic tool, Wanderer permitted them use of no other information in arriving at their final decisions. Though Hammer's criticisms are well taken, one wonders if his arguments might not fall short of offsetting the total impact of the study. When the 20 foremost experts on an assessment technique fail to do better than chance, it is difficult to see how the average clinician may be expected to do any better.

Seemingly at the heart of many of the negative findings regarding the DAP has been the general unreliability of the sign or signs being assessed (Hammer & Kaplan, 1964, 1966; Handler & Reyher, 1965; Swenson, 1968). One of the most promising principles to have emerged from these studies of drawing consistency is that the more general the type of judgment or drawing aspect being considered, the more reliable are the results obtained. Thus, for both studies of interjudge and test-retest reliability, most confidence is able to be placed in the global characteristics of the protocols (for example, level of adjustment, maturity, and degree of sexual differentiation), next most in the structural aspects (size, placement, omissions, distortions, and so on), and least of all in the specific bodily parts (such as the hands, feet, eyes, and hair). Nonetheless, it should be noted that while lower reliabilities may limit the validity of the measures involved, the higher reliabilities associated with global assessment do not assure that such assessments are valid. Grams and Rinder (1958), for example, in a study of homosexual indicators, were able to obtain interjudge agreements of roughly 80% as to the presence or absence of specific features; but the judges were

not able to differentiate successfully a normal from a homosexual group on the basis of the agreed upon indicators. These appropriate precautions notwithstanding, the general principle emerging from these studies seems a good one and suggests that, at least for the present, studies requiring clinicians to make judgments of specific DAP aspects are destined for failure.

The conclusion to be drawn from the above findings is that, with few exceptions, the numerous hypotheses set forth by Machover have not been experimentally verified. In searching for an explanation for this result, several types of answers suggest themselves. One, as proposed by Hammer (1959), is that many of the DAP pathognomic indicators are relatively rare. The result is that, while they are clinically useful, they remain statistically insignificant. Hammer notes further that since there are many types of selves, any one or combination of which may be projected in the DAP, attempts to employ a measure of only one type as a criterion are bound to lead to negative results. Though such reasoning may be clinically sound, the present investigator is inclined to agree with Murstein (1965) who questions whether such "rare" signs can really be of practical value in a clinical setting and whether the "many selves" notion can be useful when as yet there is no sound way of determining which self is being projected. In Murstein's words, Hammer seems to have changed the DAP from "a moderately complex unsubstantiated instrument to a more complex unsubstantiated instrument" (1965, p. 378).

A second explanation for the literature findings was probab-

ly first suggested by some skeptic who dismissed the DAP as nothing more than a test of art ability. Though Hammer (1958) rather summarily rejected the idea, there has developed a fair body of research which suggests that clinicians, when they think they are judging pathology, are in fact considering only the artistic quality of the drawings before them. Typically, the paradigm used in these studies is one in which drawings are first obtained from one or more groups, some judgment of artistic merit is obtained for each drawing, and finally the drawings are given to clinicians who evaluate them for level of adjustment. The first, and by now classic, study in this area was that of Whitmyre (1953) who obtained correlations of about .80 between psychologists' ratings of adjustment and artists' ratings of artistic quality. Furthermore, Whitmyre demonstrated that psychologists themselves do not judge art any differently than they judge pathology, the r between these two sets of judgments being in the high .80's. These findings were soon corroborated by Sherman (1958a, 1958b) who obtained similar results using essentially the same paradigm except that he had judges give dichotomous ratings rather than using the ranking procedure of Whitmyre which he felt not typical of clinical usage.

More recently, attempts to extend the generality of these findings have been made by several investigators. Since the Whitmyre and Sherman studies had employed global clinical judgments only, the next logical step was to determine whether the findings applied also to judgments of specific body parts. In

examining this possibility, Feldman and Hunt (1958) raised also the question as to whether pathology might in some way cause poor art. Though not testing this latter possibility directly, they reasoned that if psychic conflicts did affect art quality there was still no a priori reason why they should do so in precisely those body parts which were most difficult to draw. Asking art instructors to rate the difficulty of rendering various body parts, the authors found that clinicians' ratings of adjustment correlated $-.53$ with drawing difficulty. Similar findings have more recently been reported by White (1969) who in addition obtained the interesting, if not quite understandable, result that while artists' ratings of difficulty correlated with pathology ratings, the Ss' own ratings of drawing difficulty did not. This finding seems deserving of further investigation since it constitutes the only negative result in the literature on the subject.

Other investigators have attempted to be more specific in their study of the variables underlying the correlations between ratings of art quality and ratings of clinical judgment. Woods and Cook (1954), for example, studied placement of the hands and found this factor related to drawing proficiency, at least among their sample of eighth grade Ss. Levy, Lomax, and Minsky (1963) attempted to delimit the concept of "artistic quality" by operationally defining it in terms of proportionality of body parts. As expected, the authors found that proportional accuracy seemed to be the determining factor in a variety of clinical judgments including overall adjustment, dependency,

sexual difficulty, and intelligence. This relationship with estimates of intelligence served to confirm the earlier findings of Bieliauskas and Bristow (1959) who reported that when art majors and nonart undergraduate students, matched for ACE scores, were given the H-T-P, the art majors obtained significantly higher scores on Buck's raw G, net weighted score, good score, and flaw score. Finally, and perhaps most compelling in showing the pervasive influence of art, has been the factor analytic study of Nichols and Strumpfer (1962) which attempted to determine the dimensions present in a number of drawing scores. For this purpose, DAP protocols were scored on the basis of a number of scales devised in other studies. Among these were a sexual differentiation scale, an adjustment scale, and scales for aggression, art quality, and maturity. Though four factors emerged from the resulting analysis, one factor--described as overall art quality--accounted for most of the variance for those aspects considered clinically significant. The authors concluded that different types of clinical judgments are different in name only.

Lending support to the notion that only art quality is being judged from the DAP are a number of studies which have shown that the amount of training and clinical experience one possesses fails to aid in the making of accurate clinical judgments derived from the DAP test. Schaeffer, (1964), for example, found no differences in judgmental accuracy between groups of clinical psychologists, VA trainees, and nonpsychologists when given the task of estimating level of adjustment on the basis of DAP

protocols. Similar results have been recorded by Albee and Hamlin (1950), Simms, (1951), and Watson (1967). Recently, in fact, Stricker (1967) found a perfectly negative correlation between years of clinical experience and the accuracy obtained in sorting drawings into patient and normal groupings. The explanation was offered that whereas clinical students were willing to use the latest DAP research findings in evaluating projective drawings, the practicing clinicians were reluctant to abandon the approaches to the test which they had been employing for years. Only in one study (Guinan & Hurley, 1965) did the judgments of PhD's prove more accurate than those of less experienced persons. It is perhaps significant that in this study the effect of the PhD's greater clinical experience was, in fact, minimized as the only requirement was that drawings made by the same person on two different occasions be matched. It might be noted that these studies do not mean that a clinician cannot successfully use the DAP, since it has been shown that some psychologists can learn to use it with a satisfactory degree of accuracy (Murray & Deabler, 1958; Schmidt & McGowan, 1959). This ability to use the DAP, however, has not been shown to be in any way dependent upon psychological training; and it may be that other nonpsychologists might learn to use the test as well.

What might be concluded from these findings? Firstly, it seems significant to note that practically every study done in this area has been consistent in demonstrating the pervasive influence of artistic quality on clinical judgments derived from

the DAP. The only exception is White's (1969) finding that students' own ratings of drawing difficulty do not correlate with clinicians' estimates of adjustment. It goes without saying that in the literature on projective tests, such substantial agreement is rare indeed. The remaining problems, therefore, seem mainly ones of determining the generality of these results and the factors which work to affect the relationship between art quality and clinical judgment. This writer is aware of only a single study (Marais & Strumpfer, 1965) which has attempted to go beyond the mere establishment of this relationship. Encouragingly the investigators demonstrated that scores on the DAP body image disturbance scale (BIDS) could be validly related to body imagery if they controlled for art quality in the drawings of their experimental groups. Scores on the BIDS had previously been shown to be highly correlated with ratings of artistic quality (Nichols & Strumpfer, 1962).

Thus the question of the generality of the findings obtained in the "art-influence" studies would still seem to be open to investigation. If one examines the procedures employed in these studies, one is struck by the fact that the clinical judges are invariably forced into using somewhat artificial decision making procedures. In the first instance, every study has employed either a ranking technique or a dichotomous rating scale, neither procedure affording the clinician any opportunity to state the degree of pathology he thinks is reflected in the drawings. More important is the fact that all studies in-

sisted that the clinical judges make a decision about every drawing. In actual clinical practice, however, it seems likely that, while the DAP may be routinely administered, not every DAP is equally useful to a given clinician. Sometimes, in fact, the clinician must admit that he "gets nothing" out of a particular test. It might be added that this situation, when it arises, is not necessarily due to any inherent weakness in the test or lack of interpretative skills on the part of the diagnostician. Equally likely is the possibility that some persons reveal themselves most in projective drawings while others do so in thematic tests, in response to inkblots, and so on. At any rate, the possibility remains that when judges are forced into making decisions in the face of subjective uncertainty, they unconsciously respond to the most salient feature of the drawing (art quality) as a means of complying with the task. Giving the judges an opportunity to state the confidence they place in their judgments would seem to be a necessary addition to the usual paradigm.

Other, though less serious, objections can be made to some of the procedures used in the studies investigating the effects of art quality. Though an obvious question is whether the relationship between art quality and clinical judgment is the same for protocols from all types of Ss, more than half the studies cited used only normal or psychiatric groups (Bieliauskas & Bristow, 1959; Feldman & Hunt, 1958; Lewinsohn, 1965; White, 1969; Woods & Cook, 1954). Furthermore, Sherman's (1958a, 1958b) use of psychiatric aides as "normal" controls is, as Swenson (1968) notes, a questionable procedure at best. As for those

studies investigating the influence of art quality on clinical judgments of specific body parts (Feldman & Hunt, 1958; White, 1969), it might well be objected that the clinical judges were working, so to speak, with two strikes already against them. First, as discussed above, asking judges to deal with specific aspects of drawings is asking them to deal with what is notoriously unreliable and hence necessarily invalid. Secondly, and by way of "assistance," the judges were given copies of many of Machover's hypotheses which had, however, been shown to be of dubious value on the basis of previous research. It seems hardly surprising, therefore, that the obtained results were negative. It is noteworthy also that in these latter two studies the art judges evinced considerable disagreement over rank ordering body parts in terms of drawing difficulty and that the individual S's actual drawings were never judged for artistic merit.

Nonetheless, it is not what previous studies have done so much as what they have failed to do that seems important. As Swenson (1968) recently noted, essentially all that has been established is the fact that ratings of artistic excellence and ratings of adjustment are consistently and highly correlated. Little, however, has been done to analyze the factors affecting this relationship. Granted that art quality is a variable leading to erroneous judgment, the question remains as to whether the distracting influence of this variable can in any way be overcome. Might clinicians improve on the accuracy of their judgments

as Sherman (1958a) suggests if they limited themselves to drawings of either very good or very poor art quality? Or might Hammer's (1968) suggestion of using an extended series of drawings work to offset the art effect? To date, attempts to answer these and similar questions are not to be found in previous research.

One might summarize the results of this literature review by noting that while the DAP continues to be a popular psychodiagnostic instrument, the research findings of the past 20 years have lent support to only a few of the hypotheses associated with the test. The basic notion of projection, as expressed in the body-image hypothesis, has found some measure of support as has the notion that sexual inversion will often express itself by S drawing as his first figure a member of the opposite sex. Still, clinical application of the former finding has been hampered by the failure of past research to delimit adequately the construct validity of the body-image construct, while the sex of the first drawn figure has been shown to be a rather complicated phenomenon related, at least, to the age and sex of S as well as to any tendencies toward sexual inversion. As for Machover's many content and structural hypotheses, studies to date have led to negative or inconclusive findings. The best explanation for this conclusion seems to be found in the general unreliability of the particular DAP sign being assessed. In this regard, the literature seems conclusive in demonstrating that reliability decreases rapidly as one's consideration passes from the more general to the more specific features of DAP protocols. Nonetheless, research has shown that the increased reliability of general features does

not necessarily result in increased validity, so that, in general, clinicians have done poorly in assigning accurate diagnoses on the basis of DAP protocols alone. This last finding appears to be independent of the clinical experience of the person who does the judging.

Several explanations for these generally negative findings have been suggested in the literature. One is that the best DAP pathognomic indicators are relatively rare and thus are found to be statistically nonsignificant by the usual research paradigm. Another suggestion is that many types of self-concepts are projected into a DAP protocol, while research has relied almost exclusively on measures of the "real self" as a criterion for clinical judgment. Finally, there has been the explanation suggested by a number of studies, that clinicians, in making their clinical evaluations, are more influenced by the artistic quality of a DAP protocol than by any psychologically relevant factors which may be expressed in it. This last finding has been consistently demonstrated whether the overall art quality of the drawing or just the art quality of specific body parts was being considered. Further, art quality has been shown to be the largest single factor in accounting for the variance of a wide variety of clinical judgments, ranging from judgments of overall adjustment to those of aggression and sexual differentiation.

Generally, the methodological and statistical techniques employed in DAP research have greatly improved over the past decade (Swenson, 1968). An important shortcoming, however, would

seem to be the frequent failure of investigators to utilize the findings of other studies. From the point of view of the present study, the most serious oversights have been concerned with artistic quality, a variable which has never been controlled except in those studies specifically investigating its effects. This is difficult to understand, since control of this variable could conceivably help explain many of the contradictory research findings which abound in the literature. As for the studies of art quality themselves, many have had procedural weaknesses which have hindered attempts at generalizing their findings. Many for instance, have used only normal or patient groups, but not both; others never had Ss' actual drawings rated for art quality. Another point is that all studies have literally forced judges to make evaluations about every DAP presented to them, regardless of how sure they felt of their judgments. Finally, while noting the apparent importance of art quality in DAP evaluations, no research has been done which has attempted to see if there were any means by which the influence of art quality could be lessened.

The major differences introduced in this study were concerned more with the making of the clinical judgments than with the obtaining of art ratings. First, the study employed a 7-point rating scale instead of the usual 2-point scale typically used in previous research. It was felt this procedure would provide clinicians with a greater opportunity to express the true nature of their impressions concerning the degree of adjustment reflected in the drawings. In addition, clinicians were provided with a

6-point confidence-in-judgment scale used in conjunction with each clinical judgment. The purpose here was to remedy the forced-judgment procedures of previous studies so that clinicians would have an opportunity to state just to what extent, if any, they would actually make use of their interpretations of a given DAP in writing a clinical report. It was further felt the procedure would deal squarely with many of Hammer's criticisms of DAP research--criticisms which imply that clinicians are forced into making absolute types of judgments they ordinarily would be quite reluctant to make. Thirdly, the study explored the possibility that the influence of artistic quality might be offset: 1) by providing clinicians with an extended series of drawings as suggested by Hammer (1968); and 2) by providing clinical judges with copies of the latest research findings regarding the DAP which have met with general agreement.

Finally, it was decided to examine three general propositions concerning the relation of art quality and clinical judgment which seem not to have been examined adequately by previous research. The first was that clinical judgments are the result of an interaction between the artistic quality of a drawing and the level of adjustment of the S doing the drawing. It was noted that while studies to date have shown art to be the major factor influencing clinical judgment, the possibility of an interaction effect with adjustment level has not been ruled out.

The second proposition was that pathology results in poor art. This possibility seems not to have been directly assessed by previous research, though Feldman and Hunt (1958) tested such

a proposition indirectly. They reasoned that if pathology did act to lower the quality of artistic productions, there was no reason to suppose that it would do so in precisely those body parts which were rated by artists as most difficult to draw. Nonetheless, this is precisely what the authors found. Using the H-T-P, Lair and Trapp (1960) found no differences between psychiatric and normal groups in the artistic quality of Person drawings but did report that the House drawings of the psychiatric group were somewhat worse than those of the normal controls. This latter finding, however, was incidental to the main purpose of the study and was based on the observations of a single judge. Further, no attempt was made to determine whether a house is more or less difficult to draw than a person. The present investigation tested the proposition that pathology results in poor art by asking Ss to submit the drawing of a car in addition to the usual DAP. It has been found elsewhere (Handler & Reyher, 1964, 1966) that both drawings are equally difficult to execute but that the person drawings are more psychologically meaningful as measured by the GSR technique. The procedure has the advantage that S can provide his own control for artistic level. It was felt that if pathology does act selectively to cause poor art, then the person and car drawings of normals should receive about the same artistic quality score, whereas these scores should show some discrepancy for the psychiatric group.

Lastly, this study examined the possibility that a clinical

judge would be influenced by the art quality of the drawings to the extent that he himself could draw well or poorly. Since neither previous research nor this investigator's intuition suggested the nature of the possible relationship, no specific prediction was made as to the possible outcome.

Chapter III

Procedure

Phase I

Subjects

The Ss were 83 normals and 61 psychiatric patients. All Ss were males. The normal group consisted of students from Loyola University, Chicago, selected on a volunteer basis for testing. The psychiatric group was composed of outpatient volunteers from the VA West Side Hospital, Chicago, Illinois. Ss in the normal group ranged in age from 18 to 45 ($\underline{M} = 25.2$; $\underline{SD} = 6.82$); patient ages varied from 19 to 51 ($\underline{M} = 33.6$; $\underline{SD} = 10.19$). No attempt was made to control for age or intelligence of the groups since there is no research which has shown these factors to influence the expression of adjustment in the drawings of adults. What evidence there is suggests that intelligence ceases to be reliably related to the personality characteristics expressed in the DAP after adolescence (Harris, 1963). The larger number of Ss in the normal group was needed in order to provide a pool from which to draw in carrying out certain matching procedures described below.

Normality of adjustment was assumed to exist among the members of the student group unless S had previously sought help for psychological problems. Self-report was relied upon in obtaining this information. If psychological help had been sought, the protocol of that S was excluded from the normal sample. The protocols of two Ss were excluded for this reason.

In selecting the psychiatric group, all patient volunteers were considered eligible unless there existed a history of mental retardation or organic involvement. Hospital files were consulted to determine whether either of these conditions existed. At the time of the testing, 59% of the patients were carrying a diagnosis of psychosis, 34% were classified as neurotic, and 7% considered as characterological disorders. All were receiving outpatient treatment. Many were in a state of fair remission and holding jobs.

Procedure

Each S was requested to submit either a set of three drawings or an extended series of five drawings. Twenty patients and 20 normals did the extended series while 41 patients and 63 normals submitted the three-drawing set. A table of random numbers was used to determine which Ss from the normal and psychiatric groups would be asked to do which series of drawings. The three-drawing series consisted of a man, a woman, and a car. The five drawings composing the extended series were as follows: a man, a woman, the self, a family, and a person-in-the-rain.

Both the normals and psychiatric patients were asked to do the drawings in individual testing situations. Each S was seated comfortably at a table or a desk and given a No. 2 pencil and blank white paper 8½ by 11 inches in size. Instructions consisted of telling each S to "Draw the best man that you can," "Draw the best car you can," and so on. With the exception that Ss were told to draw whole persons, no other restrictions were put on the type of figure they might draw. Any further questions

S might have had were answered by saying "Do whatever you wish." After completion of each drawing, the finished drawing was removed and S provided with another sheet of paper.

Counterbalancing for the order of doing the drawings within the three-drawing series was achieved by the following method. Representing the man drawing by A, the woman by B, and the car by C, six possible orders are obtained in which drawings may be requested from an S (ABC, ACB, BAC, etc.). Which of these six orders was used with a given S was determined by use of a table of random numbers. The same randomized procedure was used also in determining the order of the drawings for those receiving the extended series, though only 40 of the 120 possible combinations could be employed.

Phase II

Subjects

Two male and three female Ss, all with a formal background in art, served as judges of the artistic quality of the drawings. Three of these persons were former art majors while two had pursued art mainly as an avocation but had taken courses at art schools and had themselves done several paintings for profit. None had had any experience with the DAP test.

Procedure

For several practical considerations, it was considered necessary to have more than one copy of the various drawings. To this end, five sets of copies were made on a recently serviced Xerox 914 Copier. The original drawings were not presented to any

of the judges. Each judge received copies of the drawings with the following explanation:

The Xerox machine used to copy the drawings was chosen after sampling several machines for clarity of reproduction. The machine finally selected was considered excellent in its ability to reproduce fine qualities from the original drawings. Two main features, however, served to distinguish the original from the copy: 1) the coloring; that is, the original was done in pencil gray whereas the copy is done in black ink; and 2) the shading. This shading effect was such so as to exaggerate the extremes of line pressure; that is, very light areas of the original appear lighter still on the copy, and very heavy areas of the original appear darker on the copy. These differences were slight but noticeable. Line quality of average pressure appeared equally dark on both copies. Other qualities, including erasures, were equally apparent in both copies.

All drawings collected for use in any phase of the study were presented to the art judges for evaluation. Drawings from a single person had previously been stapled together to form a set. Judges received the following instructions:

You are being given a number of sets of drawings and are asked to make a rating of overall artistic quality for each set of drawings. In some cases there will be five drawings to a set, in some two, and in some only a single drawing. In any event, all drawings from a single set should be used in arriving at a single rating of overall artistic quality.

No formal definition of art quality is being used in this study. In arriving at your final decision, you should rely on whatever notions of artistic excellence you have used in the past. In rating the drawings, the following scale will be employed:

- 1) Extremely poor.
- 2) Quite poor.
- 3) Slightly below average.
- 4) Average.
- 5) Slightly above average.
- 6) Quite good.
- 7) Excellent.

In presenting the protocols to the judges, all drawings from

the study were mixed together by assigning each a code number and using a table of random numbers. The car drawing from a given S was presented to the judges independently of the person drawings done by the same S. Since both types of drawings were to be later compared for artistic quality, this procedure was necessary to insure the independence of the two ratings. An added difficulty arose from the probability that as a judge went through the series of drawings, his notions of "average" art quality would tend to change as he saw more and more protocols. To control for these shifting frames of reference, a different random order of presentation was used for each judge.

Phase III

Subjects

Two groups of judges, a naive and an informed group, were selected to make the clinical evaluations of the drawings. Each group consisted of five male graduate students in clinical psychology currently completing their internships at various facilities in the Chicagoland area. Generally, the judges had only their dissertations to complete before receiving their doctorate. All but two had completed their course work. Average time spent in graduate school among the two groups was 4.7 years. All judges were familiar with the DAP and used it in their clinical work.

Judges were randomly assigned to either the naive or informed groups. The naive group was left to evaluate the DAP protocols solely in terms of their personal clinical experience with the test. These judges received the explanation on Xeroxing given to the art judges. They received no special instructions but

were asked to make their ratings using whatever procedures they had found useful when using the DAP in their clinical work.

The five judges in the informed group were given a sheet of research findings on the DAP, following a procedure first used by Stricker (1967). They were asked to read over and study this sheet carefully and to use it as best they could in forming their final judgments. The research findings were based mainly on the conclusions of a study by Hiler and Nesvig (1965). The form in which the findings were presented to the judges was borrowed from the article by Stricker. The introductory remarks and the last two general findings were supplied by the author. These findings with the following instructions were presented to the informed judges:

Within recent years a number of findings have emerged from the literature on the DAP which are relevant to the task which concerns you in this study. More specifically, attempts have been made to determine the basis on which accurate clinical judgments are made. In general, it has been found that the characteristics present in the drawings of the patients which aided the clinicians in judging them as such were:

- 1) Bizarreness--this category includes such impressions as "schizy," "grotesque," "inhuman," "sinister," "sick," "ghoulish," "weird," and "gnome-like" but not simply "peculiar" or "distorted".
- 2) Omissions of major parts of the body--the omission of major parts of the body such as head, body, arms, legs, hands, and torso were more characteristic of patients than of normals.
- 3) Distortion--this category was particularly effective if distortion of the head or arms was present.
- 4) Transparencies--this category referred particularly to transparency of the body or legs through the clothing.

In general, the characteristics present in the drawings of normals which aided clinicians in judging them as such were:

- 1) A happy, pleasant facial expression.
- 2) Nothing pathological--this subjective impression was much more common in the drawings of the normals.

A number of cues which, though they may be of some use in other situations, were not found useful in discriminating normals from psychiatric patients were:

- 1) Conflict and anxiety indicators--these include line emphasis, erasures, shading, nude figures, sketchiness. These qualities were not present in either group with any greater frequency.
- 2) Size and pressure--neither size nor lightness of pressure discriminated significantly.
- 3) Proportion between body parts.
- 4) Motion and posture (though an active, outgoing type of posture was slightly more frequent among normals).

Finally, two general findings have emerged from the literature which you might keep in mind when rating the drawings:

- 1) Clinicians are frequently responding only to the artistic quality of the drawing when they think in fact they are judging adjustment; that is, good art ability is frequently associated with judgments of adjustment and poor art with judgments of pathology.
- 2) A global type of analysis is more reliable and valid than an atomistic approach. It has therefore been suggested that clinicians first approach the drawings as a whole rather than taking an immediate part by part analysis.

Procedure

Judges in each group received the male and female drawings of all normal and psychiatric Ss as well as the sets of extended series drawings done by members of both adjustment groups. All drawings done by a single person were stapled together. Drawings of the automobile were not presented to the clinical judges. In presenting the drawings to the judges, all sets of drawings were randomly mixed together, again through the use of a table of random numbers. To control for serial position effects in the judging of adjustment, a different random order of presentation

was used for each judge.

Judges of each group were asked to make three ratings concerning each set of drawings. For this purpose, a 7-point adjustment scale and a 6-point confidence-in-judgment scale were devised. The choice of an even number of points on the latter scale was made since it was felt an odd number would suggest a neutral point of confidence which would be difficult to employ satisfactorily. The third rating, that of patient-nonpatient status, required no scale. All judges received the following instructions:

You are being given a number of DAP protocols and are asked to make three separate ratings for each set of drawings: 1) a rating of adjustment; 2) a rating of confidence in your judgment; and 3) a rating of patient-nonpatient status. In some cases, there will be two drawings from a given individual, in other cases, five. In either event, all drawings from one person should be used in arriving at a single rating of adjustment, a single rating of confidence, etc.

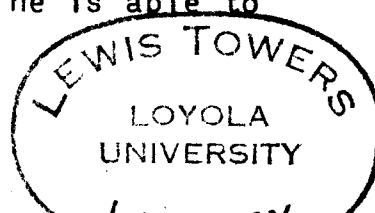
The scale to be used in making the ratings of adjustment is the following:

- 1) Extremely maladjusted.
- 2) Considerably maladjusted.
- 3) Mildly maladjusted.
- 4) Average.
- 5) Somewhat better than average.
- 6) Considerably better than average.
- 7) Extremely well adjusted.

To help clarify the meaning of the scale points, the following definitions of three of the points are given:

Scale point #1--(Extremely maladjusted.) Check this point if you feel the person is psychotic or nearly psychotic; that is, if you feel he is experiencing a serious breakdown in integrative processes.

Scale point #4--(Average.) Check this point if you feel the person has made a generally normal adjustment; that is, while having the usual number of defenses and conflicts, he is able to function reasonably well.



Scale point #7--(Extremely well adjusted.) Check this point if you feel the person approaches what has been called the "fully functioning individual"; that is, the type of person who is open to the full range of his experiences, one who is aware of and can respond to his own needs and feelings and those of the persons with whom he interacts.

The second scale is to be used for rating the confidence you have in your rating of adjustment. In this regard, it is common clinical knowledge that some test protocols are more meaningful to a given clinician than others. At times, the clinician feels almost certain of his interpretations; at other times, he feels he gets little or nothing out of a particular test. It is these extremes of clinical experience to keep in mind when using the following confidence-in-judgment scale:

- 1) Rating arbitrary.
- 2) Quite uncertain.
- 3) Somewhat uncertain.
- 4) Somewhat confident.
- 5) Reasonably confident.
- 6) Very confident.

As before, the following definitions are offered to help clarify the meaning of the scale points. Please study these thoroughly before proceeding with your ratings.

Scale point #1--(Rating arbitrary.) Check this point if you feel the adjustment rating you just made was given only because you had to check something; that is, you feel that the particular drawings in question were such that you did not know one way or the other about the adjustment status of the person.

Scale points #3 and #4--(Somewhat uncertain. Somewhat confident.) If your confidence in an adjustment rating slightly outweighs your doubts about it, check #4. If the reverse is true, check #3.

Scale point #6--(Very confident.) This does not mean that you feel 100% certain. It is realized that clinical judgments do not attain this degree of certainty. Check this point for those situations where you feel so sure of your analysis that, for example, you would have no doubt about putting the interpretation in a clinical report on the person--even if no supporting evidence could be found in any other test, in his mannerisms, and so on.

In general, in using scale points #2 through #5 think of the situation where you are writing a report. Check the lower part of the scale (#1, #2, #3) to the extent that you would feel your judgments from the DAP would need supporting evidence from other sources before you could include them in a report. To the extent you would feel these supporting sources unnecessary, check the higher end of the scale (#4, #5, #6).

Finally, please indicate whether you feel the drawing was done by a patient or a nonpatient. For this purpose, a simple dichotomous rating will suffice.

After completing all their ratings, each judge was himself asked to draw a man and a woman. Since what was desired was an index of how well each judge could draw, judges were informed that their drawings would not be clinically evaluated in order to reduce any possible suspicions or anxieties which might affect the quality of their productions. The judges' drawings were presented to the art judges who rated them along with the other drawings for artistic quality.

Phase IV

Subjects

It was considered desirable to determine whether the art judges' notions of "artistic quality" were at all congruent with the notions of "artistic quality" which might be held by the clinical judges. In answering this question, drawings were collected from 22 males who were asked to submit a drawing of "anything they wished." Since the only thing that was desired was to obtain a set of drawings which would reflect a varied range of artistic talent, it was felt that drawings of any type of object done by any type of person would suffice. No restrictions were thus put

on subject characteristics such as age, education, or other factors which might otherwise have been considered relevant.

Procedure

The Ss were seated comfortably, given a No. 2 pencil, a piece of $8\frac{1}{2}$ by 11 blank paper, and instructed to draw anything that they wished. The 22 drawings were then given to the art judges as well as to the judges in both clinical groups who rated them for artistic quality. Clinical judges received the task after completing their evaluations of adjustment. They received the same instructions and followed the same procedure as had been used in obtaining the art ratings on the DAP protocols from the art judges.

Treatment of Data

The basic data of the study were the ratings obtained from the art and clinical judges. In such a study, where judgments are made on the basis of test data, the judges are the Ss, the test data the independent variable, and the judges' ratings the dependent variable (Wanderer, 1969). This last distinction is particularly important when projective tests are used in a study, for if the test data are considered the dependent variable, many statistical and methodological difficulties arise which are avoided by use of rating scales. Thus for analysis of the ratings in this study, only a few statistical assumptions had to be met. Most basic were the assumptions that the data were continuous, tended to be normally distributed, and were collected in an unbiased manner. In using the adjustment scale, it was further assumed that there exists a unidimensional continuum of adjust-

ment. The nonparametric proportions tests used with the "patient-nonpatient" ratings made no assumptions other than those of basic probability theory. However, in using the biserial correlation with the "patient-nonpatient" category, it was assumed that underlying this dichotomy there was a continuous variable (McNemar, 1962). This assumption seemed not unreasonable since it is likely there were patients nearly well enough to leave the clinic while the normal group probably contained those at various stages of readiness for seeking professional help for their problems.

The use of the rating scale technique has long been one of the methods of choice for measuring human judgments (Guilford, 1954). The choice of a 7-point scale for the art and adjustment ratings was made since there is evidence to indicate that reliability increases steadily with an increase in scale points up to a limit of seven to nine steps (Nunnally, 1969). A 6-point confidence scale was chosen since a neutral point of confidence seemed ambiguous and might easily lead to errors of central tendency (Guilford, 1954). Full verbal descriptions were provided for various scale points as a means of anchoring the scale and to offset the tendency to avoid using the extreme scale points.

Chapter IV

Results

Reliabilities for the adjustment ratings of the naive and informed groups of clinical judges as well as for both groups combined are given in Table 1. Also shown are the reliabilities of the art judges' ratings for the artistic quality of the drawings. The intraclass correlation technique suggested by Guilford (1954) for use with rating scales was employed. The "one-rater" reliabilities determined by this method refer to what is essentially the average intercorrelation between any two judges. For all the judges of a particular group or for the average of their ratings, the "all-rater" reliabilities are provided. An inspection of Table 1 reveals that both one-rater and all-rater reliabilities for all groups of judges were highly significant ($N = 144$, $p < .001$). One-rater correlations ranged from .60 to .72 while all-rater reliabilities fell between .85 and .94.

Hypothesis I predicted that an interaction effect between artistic quality and Ss' level of adjustment would be found for the adjustment ratings assigned to the drawings by the clinical judges. Using Ss' patient-nonpatient status as a criterion of adjustment and the art ratings assigned to their drawings, four groups of 12 drawings each were selected from the total sample of drawings. Group I was thus represented by nonpatients who drew very well; group II by nonpatients who drew very poorly; group III by patients drawing very well; and group IV by patients drawing very poorly. Matching of art ratings between patients and nonpatients was undertaken for both very good and very poor

Table 1
Intraclass Reliabilities of Ratings for
Clinical and Artistic Judge Groups

Group	One rater	All raters
Clinical-naïve	.65*	.90*
Clinical-informed	.62*	.85*
Combined clinical groups	.60*	.94*
Artists	.72*	.93*

* $p < .001$.

drawing groups. The larger number of drawings in the total sample permitted the matching to be nearly exact. In only one instance was a patient drawing unable to be paired with a nonpatient drawing having an identical art quality rating.

Tables 2, 3, and 4 show the results of a 2×2 analysis of variance of adjustment ratings for both the naive and informed clinical judges and for both groups of judges combined. As expected from previous research (Sherman, 1958a; Whitmyre, 1953), the main effect for art quality was significant for all judge groups ($df = 1/44$, $p < .0001$). The hypothesis of an interaction effect between art quality and adjustment status was not confirmed for any group of judges. Unexpectedly, a significant main effect for adjustment was found for both the naive ($F = 4.633$, $df = 1/44$, $p < .05$) and the combined clinical judges ($F = 5.055$, $df = 1/44$, $p < .05$). Nonetheless, comparison of the F values for the art and adjustment main effects suggests that the artistic quality of a drawing is the factor accounting for the majority of the variance associated with clinicians' adjustment ratings.

Planned comparisons of the mean adjustment ratings for these four groups were carried out using Duncan's multiple-range test (Duncan, 1955). As Edwards (1960) has noted, Duncan's concept of protection levels is such that when more than two means are being compared, the multiple-range technique is a powerful test for the detection of real group differences but achieves such power at the risk of committing more Type I errors. In the context of this study, Duncan's test was used merely to suggest

Table 2
Analysis of Variance of Adjustment Ratings
for Naïve Clinical Judges

Source	<u>df</u>	<u>MS</u>	<u>F</u>
Art ability (A)	1	1260.750	89.402**
Adjustment (B)	1	65.334	4.633*
A X B	1	.083	.066
Within cell	44	14.102	

* $p < .05$.

** $p < .0001$

Table 3
 Analysis of Variance of Adjustment Ratings
 for Informed Clinical Judges

Source	<u>df</u>	<u>MS</u>	<u>F</u> *
Art ability (A)	1	875.521	78.234*
Adjustment (B)	1	28.521	2.549
A X B	1	25.520	2.280
Within cell	44	11.191	

* $p < .0001$.

Table 4
 Analysis of Variance of Adjustment Ratings
 for Combined Clinical Judge Groups

Source	<u>df</u>	<u>MS</u>	<u>F</u>
Art ability (A)	1	4218.750	115.858**
Adjustment (B)	1	184.083	5.055*
A X B	1	27.000	.741
Within cell	44	36.413	

* $p < .05$.

** $p < .0001$.

implications for future research. The results of comparing each of the four group means with each other revealed that for both the informed and the combined judge groups all comparisons were significant ($p < .05$) except the comparison of adjustment ratings between the patient and nonpatient good-art groups. Similar findings were obtained for the naive judges except that comparison of means for the patient and nonpatient poor-art groups was also nonsignificant. Of interest here is the fact that while clinicians' adjustment ratings failed to differentiate between patient and nonpatient groups who drew well, among the poor-art groups patient drawings received significantly lower adjustment scores.

To further evaluate the role of artistic quality, Pearson r 's were calculated between each judge's adjustment ratings and the mean art rating assigned to each drawing. Correlations ranged from .42 to .73 with an average r of .62 for the ten judges combined. All r 's were significant ($N = 144$, $p < .0001$). Average correlations for the naive and for the informed groups of judges were .66 and .58 respectively, the difference between these two r 's being significant ($p < .05$, two-tailed test). The results point to the conclusion that clinicians' ratings of adjustment are strongly related to whatever art judges feel is good or poor artistic quality.

In attempting to go beyond the literal interpretation of the obtained r 's, the question arose regarding the likelihood of the art judges' idea of fine art being congruent with that of the

clinicians'. To answer this question, the 22 drawings of various objects obtained during phase IV of the study were employed. Comparison was then made between the art judges' ratings and the art ratings assigned by the clinical judges to the same set of drawings. The obtained Pearson r of .95 ($p < .0001$) suggests that the clinicians would have assigned essentially the same art ratings as the art judges to the DAP protocols had they been requested to do so.

Hypothesis II stated that the state of pathology will result in poor art to the extent the object drawn is personally meaningful to the individual. To examine this notion, mean art ratings were obtained for both the person and car drawings from both the patient and nonpatient groups. Previous studies by Handler and Reyher (1964, 1966) had shown that both types of drawing were equally difficult to execute but that the person drawings contained more personal meanings for the Ss involved. It was thus expected that while art ratings for both drawings from the nonpatient group should be roughly equivalent, the patient group might display greater artistic skill in their execution of the auto drawing. Testing of this hypothesis was carried out by testing the significance of the difference between difference scores; that is, the mean difference (\underline{M}_{D_1}) between the nonpatients' car and person art ratings was compared with the mean difference (\underline{M}_{D_2}) between the patients' car and person art ratings. A two-tailed test of significance between the two \underline{M}_D 's failed to support the hypothesis that pathology results in

poor art for objects containing much personal meaning ($M_{D_1} = .52$, $SD_{D_1} = 5.12$; $M_{D_2} = .29$, $SD_{D_2} = 4.55$; $p > .10$). For the nonpatient ss the Pearson r between art ratings for the car and the person drawings was .65; for the patient group, $r = .81$. Fisher's r to z transformation, employed in testing the significance of the discrepancy between these two r 's, resulted in a nonsignificant difference ($p > .10$).

Hypothesis III stated that providing the clinician with research findings on the DAP would increase the accuracy of his judgments. This hypothesis was tested by comparing the proportions of correct patient-nonpatient classifications for the naive and the informed clinical judges. For the naive group, the percentage of correct classifications was 62.8%; for the informed group, 64.2%. A test of the significance of the difference between the two proportions failed to support the hypothesis ($p > .10$).

Hypothesis IV was that providing the clinician with an extended series of drawings would increase the accuracy of his judgments. To control for the effect of artistic quality, the 20 patient and 20 nonpatient drawings of the extended series were matched on the basis of art ratings with equal numbers of patient-nonpatient drawings from the two-drawing protocols. Due to the larger number of two-drawing sets, matching on the art ratings was exact. In cases where more than one of the two-drawing sets might have provided a match for a single extended set, a table of random numbers was used to select the protocol to be used in the

final analysis.

Table 5 shows the proportion of correct patient-nonpatient judgments for the two-drawing and five-drawing series. Significance tests for the difference between the proportions of both series of drawings were carried out for each judge independently, for both the naive and informed groups considered separately, and for both clinical judge groups combined. Examination of Table 5 reveals that the hypothesis of an extended drawing series increasing accuracy of judgment was supported for most comparisons made. Results were significant for five of the ten judges considered singly ($p < .05$ to $p < .01$), for the total judgments of the naive ($p < .0001$) and informed ($p < .01$) groups, and for both groups of clinical judges combined ($p < .0001$). Only in the case of a single judge were judgments based on the extended series less valid.

Factors relating to a judge's expressed degree of confidence in his adjustment ratings are reported in Tables 6 and 7. All analyses were carried out for each of the judges individually, for each group of judges separately, and for both clinical judge groups combined. Column 1 of Table 6 presents biserial correlations between the judges' degree of confidence for a drawing and their assignment of patient or nonpatient status to the drawing. In most cases, a significant relationship between heightened confidence and assignment of patient status was found. The r_b 's were significant for five of the ten judges considered singly (significant r_b 's ranged from .21 to .76, $p < .01$ to $p < .001$), for

Table 5
Proportion of Correct Patient-Nonpatient Judgments
for Two-Drawing and Five-Drawing Series

Rater	Five drawings	Two drawings	$p_1 - p_2$	z
Naïve judges				
Judge A	.800	.600	.200	1.75*
Judge B	.650	.550	.100	.87
Judge C	.750	.450	.300	2.59**
Judge D	.750	.500	.250	2.41**
Judge E	.750	.575	.175	1.38
All judges	.740	.535	.205	4.61***
Informed judges				
Judge F	.600	.500	.100	.48
Judge G	.850	.650	.200	2.02*
Judge H	.650	.500	.150	1.25
Judge I	.800	.525	.275	2.43**
Judge J	.500	.550	-.050	-.43
All judges	.680	.545	.135	2.84**
Combined clinical judges	.710	.540	.170	5.25***

* $p < .05$, one-tailed test.

** $p < .01$, one-tailed test.

*** $p < .0001$, one-tailed test.

Table 6

Correlations of Judges' Degree of Confidence with Assignment
of Patient-Nonpatient Status and Correctness of Judgment

Group	Patient status		Correct judgments	
	r_b	z^a	r_b	z^b
Naive judges				
Judge A	.14	1.33	.01	.09
Judge B	.76	6.83**	.10	.93
Judge C	-.04	.37	.30	2.86*
Judge D	.21	2.00*	.20	1.84*
Judge E	.14	1.33	.00	.00
All judges	.06	1.26	.16	3.33**
Informed judges				
Judge F	.34	3.29*	.15	1.43
Judge G	.38	3.62*	.38	3.38**
Judge H	-.01	.09	-.10	.94
Judge I	.37	3.52*	.23	2.13*
Judge J	.14	1.33	.20	1.90*
All judges	.36	7.66**	.24	4.99**
Combined clinical judges	.22	6.67**	.24	7.06**

^aTwo-tailed test.

^bOne-tailed test.

* $p < .05$.

** $p < .001$.

Table 7

F Values and Correlation Ratios of Judges' Confidence Ratings
for Drawings from Four Levels of Artistic Quality

Group	<u>df</u>	<u>MS</u> _{bg}	<u>MS</u> _{wg}	<u>F</u>	<u>eta</u>
Naive judges					
Judge A	3/76	7.046	1.200	5.872*	.43*
Judge B	3/76	2.379	.235	10.123**	.53**
Judge C	3/76	3.729	.316	11.801**	.47**
Judge D	3/76	3.033	.370	8.197**	.49**
Judge E	3/76	.846	.785	1.078	.20
All judges	3/76	57.359	8.356	6.867**	.46**
Informed judges					
Judge F	3/76	5.700	.896	6.363**	.45**
Judge G	3/76	5.056	1.019	4.962*	.40*
Judge H	3/76	1.246	.664	1.877	.28
Judge I	3/76	3.046	.438	6.954**	.46**
Judge J	3/76	2.483	.817	3.039	.33
All judges	3/76	66.146	9.001	7.238**	.47**
Combined clinical judges	3/76	244.971	20.710	11.830**	.56**

Note.--A simple randomized design was used with four levels of art ability: very good, high average, low average, and very poor; N = 20 for each group.

* $p < .01$.

** $p < .001$.

the informed judge group ($r_b = .36$, $p < .001$), and for both clinical groups combined ($r_b = .22$, $p < .001$). It would thus appear that, at least on the DAP, clinicians feel more confident in expressing a judgment of pathology than in expressing one of essential normality.

Hypothesis V stated that there would exist a positive relationship between a clinician's confidence in his judgments on the DAP and the correctness of those judgments. The hypothesis was tested by correlating the judges' degree of confidence for a drawing with the correctness of their patient-nonpatient classifications. Column 2 of Table 6 shows the results of these analyses. Examination of the biserial correlations reveals that in most cases the hypothesis was supported. The obtained r_b 's were significant for five of the ten judges considered singly ($N = 144$ for all judges, significant r_b 's ranged from .20 to .38, $p < .05$ to $p < .001$), for both the naive ($N = 700$, $r_b = .16$, $p < .05$) and informed ($N = 700$, $r_b = .24$, $p < .001$) judge groups, and for both clinical groups combined ($N = 1400$, $r_b = .24$, $p < .001$). It thus appears that a judge's feeling of confidence concerning his DAP evaluations does tend to be related to the actual correctness of those evaluations. Nonetheless, the magnitude of the r_b 's suggests that knowledge of this finding would be of little value to the average judge in making clinical decisions from a single DAP protocol.

It was felt that perhaps a judge's feelings of confidence in a judgment might be more strongly related to the degree of

artistic quality found in a drawing. To test this proposition, 20 DAP protocols were selected from each of four levels of artistic quality. Art ratings were the only criterion used in the grouping of the drawings which were selected without regard to patient or nonpatient status. The four groups were chosen as follows: the 20 drawings receiving the highest art ratings comprised the "very good" group; the 20 drawings with the lowest ratings formed the "very poor" group; and the 20 drawings on each side of the median art rating were chosen for the "high average" and "low average" groups. Confidence ratings expressed for drawings in each of these four groups were then compared. An inspection of the scattergram of the confidence ratings revealed that the assumption of linearity was not tenable. Data were thus examined by means of correlation ratios and a simple randomized analysis of variance design, following the procedures outlined by Bruning and Kintz (1968).

Table 7 presents the \underline{F} values and correlation ratios of the judges' confidence ratings for drawings from the four levels of artistic quality. It appears evident from these data that confidence ratings do vary from one level of artistic quality to another. Significant \underline{F} values were obtained for seven of the ten judges considered individually (\underline{F} 's ranged from 4.962 to 11.801, $\underline{df} = 3/76$, $\underline{p} < .01$ to $\underline{p} < .001$), for both the naive ($\underline{F} = 6.867$, $\underline{df} = 3/76$, $\underline{p} < .001$) and informed ($\underline{F} = 7.238$, $\underline{df} = 3/76$, $\underline{p} < .001$) judge groups, and for both clinical judge groups combined ($\underline{F} = 11.830$, $\underline{df} = 3/76$, $\underline{p} < .001$). Inspection of the con-

confidence rating totals for each group revealed that most confidence was expressed in the adjustment ratings of very poor drawings and next most in very good drawings, with both high- and low-average art ability groups being nearly equal and receiving the lowest confidence ratings. The strength of this curvilinear relationship was determined by calculating correlation ratios for each of the judges and judge groups. The number of significant eta's, along with their p values, was of course the same as that for the F values. Values of the significant eta's ranged from .40 to .53 for the individual judges. For the naive judge group, the correlation ratio was .46; for the informed judges, .47; and for both judges groups combined, .56.

Sherman (1958a) has suggested that clinicians might profit more from the DAP if they confined their judgments only to drawings of intermediate art quality. This possibility was evaluated by again dividing the drawings on the basis of art ratings into three groups: a good art group (N = 24), a poor art group (N = 24), and a group representing average art ability (N = 36). An equal number of patients and nonpatients, matched for art ratings, was selected for each of the three groups. Because of the large number of protocols collected, the matching on art ratings was nearly perfect. In the five instances where an exact match could not be obtained, the difference in the art scores for both members of the matched pair was only one point. Matching requirements and the limited number of good art drawings resulted in the smaller N's for the good and poor art groups.

Results of comparing clinical judgments for each of the three levels of art ability are shown in Tables 8, 9, 10, and 11. Table 8 contains the intraclass reliabilities of adjustment ratings for drawings from each of the three levels of art ability. Reliabilities were determined for both the naive and informed judge groups and for both judge groups combined. An inspection of the results reveals that ratings of adjustment were generally more reliable when they were made for drawings of poor art quality. Ratings for drawings of average art quality were intermediate in reliability, and the ratings of the good art group were the least reliable of the three. Results of testing for the significance of these differences are reported in Table 9. For both clinical judge groups combined, one-rater and all-rater reliability differences between the three art groups were significant for all comparisons made ($p < .0001$). For the informed judges, all comparisons of reliabilities were likewise significant (significance varied from $p < .05$ to $p < .0001$) with the exception of the difference between one-rater r 's for the average art and the good art groups. No significant reliability differences were found for adjustment ratings from the naive judge group.

The judges' proportions of correct patient-nonpatient judgments were tested against chance expectancy (50% correct) for each of the three art levels. Table 10 reports the results of this analysis. Though in the large majority of cases, the obtained proportion of correct judgments did exceed 50%, only in

Table 8
Intraclass Reliability of Adjustment Ratings for
Three Levels of Artistic Quality: Poor
Art, Average Art, and Good Art

Raters	Poor art		Average art		Good art	
	r_{11}	r_{kk}	r_{11}	r_{kk}	r_{11}	r_{kk}
Naive judges	.65	.90	.52	.85	.55	.86
Informed judges	.61	.89	.41	.78	.32	.70
Combined clinical judges	.59	.94	.46	.87	.21	.73

Note.--Reliability for a single rater = r_{11} ; for all raters = r_{kk} . Significance of r 's was not determined since interest lay in differences between r 's for different art levels (see Table 9).

Table 9

Comparison of Intraclass Reliabilities of Adjustment Ratings
for Three Levels of Artistic Quality: Poor Art
(P.A.), Average Art (A.A.), and Good Art (G.A.)

Rater	P.A. vs A.A.		P.A. vs G.A.		A.A. vs G.A.	
	r_{PA-AA}	\underline{z}	r_{PA-GA}	\underline{z}	r_{AA-GA}	\underline{z}
Naïve judges						
One rater	.13	1.48	.10	1.00	-.03	-.34
All raters	.05	1.32	.04	.98	-.01	-.26
Informed judges						
One rater	.20	2.27*	.29	2.90**	.09	1.02
All raters	.11	2.89**	.19	4.63***	.08	2.11*
Combined clinical judges						
One rater	.13	5.00***	.38	8.26***	.25	9.62***
All raters	.07	5.60***	.21	15.00***	.14	11.20***

* $p < .05$, two-tailed test.

** $p < .01$, two-tailed test.

*** $p < .0001$, two-tailed test.

Table 10

Significance Tests for the Proportion of Correct Patient-
Nonpatient Judgments for Three Levels of Artistic
Quality: Poor Art, Average Art, and Good Art

Group	Poor art		Average art		Good art	
	$P_{ob}-P_{.50}$	Z	$P_{ob}-P_{.50}$	Z	$P_{ob}-P_{.50}$	Z
Naive judges						
Judge A	.083	.81	.000	.00	.042	.41
Judge B	.083	.81	.000	.00	.000	.00
Judge C	.042	.41	.028	.34	.083	.81
Judge D	.042	.41	-.028	-.34	.042	.41
Judge E	.125	1.23	.083	1.00	.000	.00
All judges	.075	1.63	.017	.45	.033	.96
Informed judges						
Judge F	.083	.81	-.083	-1.00	.083	.81
Judge G	.167	1.64	.083	1.00	.208	2.04*
Judge H	.083	.81	-.028	-.34	-.083	-.81
Judge I	.125	1.23	-.056	-.67	.042	.41
Judge J	.042	.41	.083	1.00	.000	.00
All judges	.100	2.17*	.000	.00	.050	1.09
Combined clinical judges	.088	2.75**	.008	.30	.042	1.32

* $p < .05$, two-tailed test.

** $p < .01$, two-tailed test.

Table 11

Comparison of Proportion of Correct Patient-Nonpatient
Judgments for Three Levels of Artistic Quality: Poor
Art (P.A.), Average Art (A.A.), and Good Art (G.A.)

Group	P.A. vs A.A.		P.A. vs G.A.		G.A. vs A.A.	
	$p_1 - p_2$	z	$p_1 - p_2$	z	$p_1 - p_2$	z
Naïve judges						
Judge A	.083	.91	.041	.41	.042	.47
Judge B	.083	.91	.083	.82	.000	.00
Judge C	.014	.15	-.041	-.41	.055	.60
Judge D	.070	.77	.000	.00	.070	.77
Judge E	.042	.47	.125	1.25	-.083	-.91
All judges	.058	1.41	.042	.93	.016	.40
Informed judges						
Judge F	.166	1.84	.000	.00	.166	1.84
Judge G	.084	.99	-.041	-.45	.125	1.44
Judge H	.111	1.22	.166	1.66	-.055	-.60
Judge I	.181	2.01*	.083	.83	.098	1.08
Judge J	-.041	-.45	.042	.42	-.083	-.91
All judges	.100	2.50*	.050	1.11	.050	1.22
Combined clinical judges	.080	2.79**	.046	1.44	.034	1.17

* $p < .05$, two-tailed test.

** $p < .01$, two-tailed test.

one instance did the proportion correct for an individual judge permit rejection of the null hypothesis. Essentially similar results were obtained for the naive and informed judge groups and for both judge groups combined. Only in the judging of poor art drawings was statistical significance obtained. For the DAP protocols of this group, the proportion of correct patient-nonpatient judgments was 60% for the informed judge group ($p < .05$) and 58.8% for both judge groups combined ($p < .01$). Though the naive judge group did best in judging poor art drawings getting 57.5% correct, the proportion correct was not statistically significant ($p = .10$).

Table 11 shows the results of comparing the proportion of correct patient-nonpatient judgments made for one level of art quality with those of another level. It was felt that though the proportion of correct judgments did not for the most part differ from chance for the three art levels, they might differ significantly from one another. Data analysis revealed that this was generally not the case, most such comparisons failing to reach statistical significance. Of significance only was the finding that, for the informed judge group ($p < .05$) and for both judge groups combined ($p < .01$), poor art drawings were judged more accurately than those of average art quality. This difference was 10% in the case of the informed judges and 8% for all judges combined. Naive judges failed to show a significant improvement in accuracy from one level of art ability to the other. It thus appears that with the possible exception of drawings which

are quite poor in artistic quality, clinical judgments do not tend to be more accurate for drawings of one level of proficiency as compared to those of another.

Finally, the study attempted to see if the extent to which a judge was influenced by artistic quality would be a factor relating to certain other variables which suggested themselves as possibly being important. It was decided that the previously determined correlations between the adjustment ratings of a given judge and the art ratings assigned to the drawings would serve as an index of the extent to which each judge was influenced by the artistic quality of the drawings. It was first asked if a judge's own artistic ability might relate to the degree he was influenced in his adjustment ratings by the art quality of the drawings. In testing this proposition, the clinical judges were ranked on each of two variables: first, in terms of the art ratings they received on their own DAP's; and second, in terms of the magnitudes of the correlations between the adjustment and art ratings. A rank order correlation (rho) of $-.14$ was obtained but was not significant ($t = 0.40$, $df = 8$, $p > .20$, two-tailed test). The direction of the rankings on the two variables was such that the obtained rho had suggested that the better a judge was able to draw, the less he was influenced by the artistic quality of a drawing.

A second question was whether a judge's decisions regarding patient-nonpatient status were more accurate to the extent he was not influenced by the quality of art displayed in the draw-

ings. Judges were again ranked on each of two variables: first, in terms of the accuracy of their judgments concerning patient-nonpatient status; and second, in terms of the extent they were influenced by the artistic quality of the drawings. The rank correlation (ρ) between these variables of .64 was not significant ($t = 2.26$, $df = 8$, $.05 > p < .10$, two-tailed test). The direction of the correlation was, however, surprising and suggested that a judge tended to be more accurate in his judgments to the extent he was influenced by the art quality of the drawing involved.

A final analysis was concerned with whether the improved accuracy of the patient-nonpatient judgments shown for the extended drawing protocols might have been greatest for those clinicians who were most influenced by the art quality displayed in the shorter drawing series. Again, judges were ranked in terms of two variables. The z scores reported in Table 5 were used as an index for ranking judges according to how much each profited from use of extended drawing protocols. Judges were also ranked in terms of the extent to which they had been influenced by the artistic quality of the drawings. The rank correlation (ρ) of .47 failed to reach significance ($t = 1.51$, $df = 8$, $.10 > p < .15$, two-tailed test). Inspection of the data, however, revealed that the ranks of one of the judges were almost reversed for the two variables; that is, while ranked second in being influenced by art quality, he was ranked eighth in terms of profiting from an extended series. Further inspection

of the data revealed that this judge had displayed a stronger bias than any other in the sense that more than 74% of all his judgments were "nonpatient" evaluations. Further still, the same judge had given the highest average confidence ratings of all judges, seldom expressing more doubt than certainty about his ratings on any drawing. Thus, since this part of the data analysis was exploratory in nature, it was decided to recalculate the rank correlation excluding this judge from the analysis. The resulting rho of .78 was significant in spite of the loss of one degree of freedom (t = 3.33, df = 7, p < .02, two-tailed test). It thus appears probable that an extended drawing series improves the accuracy of a clinician's judgments to the extent that his judgments tend to be associated with the artistic quality of the drawings being evaluated.

Chapter V

Discussion

A survey of the literature on the DAP revealed: 1) that the DAP continues to be a popular tool in the armamentarium of the clinical psychologist: 2) that in spite of its clinical popularity, research findings of the past 20 years have offered little consistent support for the large majority of Machover's hypotheses; 3) that the most consistent finding gleaned from the literature suggests that clinicians, in making their clinical evaluations, appear to be more influenced by the artistic quality of a DAP protocol than by any psychologically relevant factors which may be expressed in it; and 4) that to date little research effort has been spent in an attempt to go beyond the mere establishment of the correlation between measures of artistic quality and the clinicians' evaluations of adjustment. It was thus the purpose of this study to investigate more fully the relationship between artistic quality and clinical judgment and to attempt to determine if there were any means by which the influence of art quality could be lessened.

The results of this study confirmed the findings of earlier investigations (Feldman & Hunt, 1958; Sherman, 1958a; Whitmyre, 1953) which have shown that the artistic quality of a drawing is the largest single factor accounting for the adjustment ratings assigned to DAP's by clinicians. The correlations between the art and the adjustment ratings in this study ranged from .42 to .73, the average r for all clinical judges being .62. These

correlations, coupled with the finding that clinicians rate art quality almost exactly as do the art judges themselves ($r = .95$, $p < .001$) would seem also to serve as an indirect confirmation of Whitmyre's (1953) finding that clinicians do not rate adjustment significantly different from the way they rate art.

In spite of these findings, it seems unwarranted (and unjustified) to conclude that the clinical user of the DAP is responding only to the artistic quality of the drawings. Results of this investigation point strongly to the fact that the issues involved cannot be so simply stated. For example, a rho of .64 was obtained between the degree a judge was influenced by art quality and the correctness of his patient-nonpatient judgments. Though this correlation did not quite reach significance ($.05 > p < .10$), it was in a direction opposite to that which would be reasonably expected and suggested that the more a judge responded to artistic quality, the more accurate were his judgments! Further support for this notion that art quality is not the only factor involved in clinical judgment was shown by the fact that while the informed judges, given the findings about artistic influence, assigned adjustment ratings based less on art quality than those of the naive judges ($p < .02$), the accuracy of their patient-nonpatient judgments was no greater ($p > .10$). Finally, the analysis of variance of the adjustment ratings for all judges resulted in a significant F value for the adjustment factor ($F = 5.055$, $df = 1/44$, $p < .05$), though the F value for the art factor was admittedly quite larger ($F = 115.858$, $df = 1/44$, $p < .0001$). These

findings, while not minimizing the pervasive influence of art quality, do lend a measure of support to the persistent belief on the part of clinicians that something psychologically relevant is being expressed in the DAP.

The hypothesis that the effect of artistic quality on adjustment ratings would vary for different levels of psychological adjustment led to less clear cut results. The main test of this hypothesis, the Art X Adjustment interaction, resulted in an F ratio which failed to reach significance. However, planned comparisons, by means of the less stringent Duncan's multiple-range test, revealed that while there was no difference in adjustment ratings between patients and nonpatients who drew very well, the mean rating for patients was significantly less among those groups drawing quite poorly ($p < .05$). This result was obtained for the informed judges as well as for both clinical judge groups combined. Further evidence for the differential effects of different degrees of artistic talent came from the finding that the reliability of adjustment ratings varied for three levels of art quality. Considering all clinical judges as a single group, it was found that adjustment ratings were most reliable for drawings of poor art quality and least reliable for those of good quality, with reliabilities for the average art group falling in between. Similar differences were found for the validity of patient-nonpatient judgments, in this case judgments being more accurate for the poor art group ($p < .01$) than for either the average or good art groups which failed to differ significantly.

Considering all these findings together, it would seem that clinical evaluations on the DAP show more consistency and greater validity if the drawings involved are artistically quite poor. Such findings fail to confirm Sherman's (1958a) suggestion that psychologists would do better if they confined themselves to drawings of intermediate artistic quality. Why judgments of artistically poor drawings should show the obtained differences is somewhat of a mystery. One possibility which suggests itself is that the bizarre elements of psychotic drawings--when they occur--are seldom associated with drawings of good or even average artistic quality. There may, in short, be a distinct subgrouping of bizarre-element drawings within the poor art group which clinicians have little difficulty in recognizing as having come from psychotic persons. Put somewhat differently, it may be that there is a level of psychosis at which pathology results in poor art. It seems not unlikely, for example, that many schizophrenics, disrupted as they are in their cognitive and perceptual-motor functioning, might also show an inability to execute a drawing commensurate with their premorbid level of artistic ability. The poor art which would result might further be expected to reflect a bizarre quality since it resulted, at least partially, from a psychiatric disturbance and not merely from an inherent lack of drawing skill. Reasoning along these lines, however, should proceed cautiously, since the findings of this study which resulted from comparing person and car drawings for two adjustment groups indicate that pathology does not result in poor art in

personally meaningful types of drawings--at least not when all degrees of pathology and all degrees of art ability are considered.

The attempts of this study to isolate factors relating to improved accuracy of clinical judgment met with varied degrees of success. As already discussed, the notion that certain levels of artistic quality might tend to be associated with greater judgmental accuracy was supported by the finding that both the reliability and validity of clinicians' ratings were best for artistically poor drawings. On the negative side, the attempt to improve the accuracy of clinical evaluations by providing clinicians with DAP research findings resulted in no appreciable gain in the number of correct judgments. This finding suggests that perhaps research results by themselves are of little value to the user of the DAP. While Stricker (1967) implied that such findings were used more effectively by Ss of limited clinical experience, his study failed to use as control groups Ss from varied levels of experience who received no research findings at all. It may be that research findings, to be useful, must be used in conjunction with the opportunity to judge and receive feedback on actual DAP protocols. Where this approach has been employed (Murray & Deabler, 1959), it has worked well, even in the absence of providing Ss with specific conclusions from the literature.

The use of an extended series of drawings, as suggested by Hammer (1968), resulted in strong support for the notion that

clinicians can significantly improve on the accuracy of their DAP evaluations by requesting their patients to submit more than the usual two drawings. Other findings of this study suggested that the effectiveness of the extended series lay in its ability to offset the tendency of clinicians to respond to the artistic quality of DAP protocols. There was, in fact, a strong positive correlation ($\rho = .78$) suggesting that those most susceptible to the effects of art quality were those profiting most from use of an extended drawing series. However, this latter finding depended on excluding one of the judges from the data analysis and therefore should be viewed with caution until the issue can be further researched. By way of further implications for later investigations, the data on the judge excluded from the analysis suggested that an extended drawing series might be of little value to the clinician who is biased toward giving set types of judgments or is overly self-assured about his ability to effectively use the DAP.

The hypothesis that there would be a positive correlation between a clinician's confidence in his judgments and the correctness of those judgments was generally supported for most comparisons made. Of greater significance seems the fact that clinicians' feelings of confidence related just as strongly to their assignment of patient status to a drawing, and related more strongly still to a DAP's level of artistic quality. Considering all ten judges as a group, the r_b of confidence ratings with correct clinical judgments was .24 ($p < .001$), while for

the relationship of confidence to the assignment of patient status, $r_b = .22$ ($p < .001$). As for the effect of art quality on a judge's feelings of confidence, it was found that the relationship between these two variables was curvilinear in nature, heightened feelings of confidence being most strongly associated with drawings having very good or very poor artistic ratings ($\eta^2 = .56$, $p < .001$).

Such findings point strongly to the conclusion that any feelings of certainty a clinician has about his DAP evaluations can be accounted for by factors having little or no relationship with the S's actual level of personality functioning. Moreover, the findings obtained in regard to a clinician's feelings of confidence would seem to undercut many of the arguments Hammer (1959, 1968, 1969) has offered by way of defending the DAP in the face of many negative research findings. Hammer has noted, for example, that the rarity of certain DAP signs precludes the possibility of their attaining statistical significance, that clinicians are asked to make decisions about each and every DAP regardless of their feelings about a particular protocol, and that judges are often forced into evaluating persons solely on the basis of the DAP when in practice they would use other tests as well. To this investigator, these points are well taken and seem essentially to imply that clinicians frequently find themselves involved in very artificial types of decision making situations in order to meet various methodological requirements. All this is true enough. Yet when such conditions exist, it

would seem that clinicians themselves, as well as Hammer, should be aware of them and able to express this awareness by means of confidence ratings which relate more strongly to correctness of judgment. The fact that confidence was far more strongly related to artistic quality implies that clinicians have somehow been led into believing they could effectively use an instrument which, for the most part, results in little better than chance-level accuracy. The only other explanation that suggests itself is that the clinical judges did, in fact, feel very unsure of their judgments but for various personal reasons were unwilling to admit this fact. If this were true, however, one wonders if the same personal reasons might not lead to the same end results under the pressures "to produce" found in many clinical settings.

Still, such considerations can rightly be applied only to the two-drawing DAP as it has typically been used in clinical practice. As already noted, Hammer's suggestion to use a larger number of drawings does appear to possess considerable merit. While this method does result in greater clinical accuracy, the time for the average S to do five drawings is not a good deal less than that required for many other tests generally considered much longer. The question thus becomes one of whether the extra time involved is worth the effort. This may well be the case. It seems likely, for instance, that some clinicians might be able to derive more from drawings than they can from other assessment techniques, or that some patients and/or personality attributes might express themselves more clearly in drawings

than in inkblots of thematic tests. Hopefully, such questions will be answered by future research.

Other considerations which seem important for future research were pointed to by this study. Not all of these were new. The findings regarding the influence of art quality, for example, merely add one more reference to the growing list of studies which have consistently shown the importance of this factor. By this time it seems obvious that future studies on the DAP must take this factor into account. This could be accomplished either by controlling for art quality in the design of the research or by making clear how the experimental manipulations of the study would affect or leave unaffected the influence of this variable. The present investigation, for example, suggested that the benefits of an extended series of drawings were greatest for those clinicians who were most affected by a drawing's artistic quality. This finding admittedly requires cross-validation in a study using a larger number of judges than employed in this investigation. Nonetheless, independent confirmation of this result would be encouraging inasmuch as it would suggest that other factors might later be discovered which would help offset the effects of art quality even more.

Not surprisingly, this investigation also pointed to the importance of treating data in terms of individual as well as group analysis. Examination of the tables contained in this report makes clear that techniques which might be employed to improve the accuracy of one judge's ratings would leave rela-

tively unaffected the evaluations of another judge. Some judges, for instance, did best with artistically poor drawings, others with those of good art quality. For some, use of an extended drawing series proved of great benefit, while others did just as well with the shorter two-drawing protocols. From this it seems probable that no single technique or set of techniques will ever be found which will be of equal benefit to all clinicians. Though global analyses seem essential in delimiting a set of possibly relevant aids to clinical judgment, it appears likely that specific sets of techniques will have to be worked out for each user of the test.

There are obviously many implications here for training practices which might be employed with the veteran as well as the inexperienced user of projective tests. One interpretation of the positive correlations of confidence ratings with judgments of pathology which were obtained in this study is that clinicians simply do not have a clear idea of how normality reflects itself in a DAP protocol. In support of this notion is the finding of Hiler and Nesvig (1965) that a clinician's best indicator of a normal DAP was that it contained "nothing pathological." A more positive conception of a normal test record might result if users of projective tests were permitted to spend more time testing, and studying the protocols of, persons who are not patients. Unfortunately, many practical concerns resulting from shortages of time and personnel make implementation of any such individually oriented training program tremen-

dously difficult to achieve.

Chapter VI

Summary

A survey of the literature on the DAP led to the conclusion that while many studies had demonstrated a strong relationship between the artistic quality of a drawing and clinicians' ratings of adjustment, little was known about what factors operated to cause this relationship or what techniques, if any, could be devised to offset it.

In an attempt to determine some of the variables operative in this area, man, woman, and car drawings were obtained from 63 male students from Loyola University, Chicago, and from 41 male outpatients from a Chicagoland VA hospital. Another group-
ing of 20 students and 20 patients, drawn from the same sources, was asked to submit an extended series of five drawings: a man, a woman, the self, a family, and a person-in-the-rain. Randomization techniques were used for varying the orders in which the drawings were done as well as for determining whether an S would do the three- or five-drawing series.

All drawings were then rated on a 7-point scale for artistic quality by five persons having a formal background in the study of art. Car drawings from an S were presented to the judges independently of other drawings done by the same person. Judges were left to form their own definitions of "art quality" in rating the drawings. To avoid the possibility of serial position effects influencing the art ratings, a different random order of presentation was used in giving the drawings to each judge.

Ten PhD graduate clinicians near completion of their training were randomly assigned to either a naive or informed group of judges. The naive group was left to rate the drawings solely in terms of their past clinical experience. The informed group received a list of generally agreed upon DAP research findings to aid them in evaluating the drawings. Both groups then rated the drawings on a 7-point adjustment scale and indicated whether they felt the protocols to have come from a patient or nonpatient. They further indicated how certain they felt of their evaluations by means of a 6-point confidence-in-judgment scale. A different random order of presentation was again used in presenting the protocols to each judge. Car drawings were not seen by the clinicians. After completion of their adjustment evaluations, the clinical judges themselves were asked to submit drawings of a man and a woman. They were finally requested to rate for artistic quality 22 drawings of various objects which had previously been collected. The clinicians' own DAP's as well as these 22 drawings were also rated by the art judges.

The study tested the following hypotheses: (a) that clinical judgments of pathology are the result of an interaction between Ss' level of adjustment and the artistic quality of their drawings; (b) that pathology results in poor art to the extent that the object drawn is personally meaningful to the individual; (c) that providing clinicians with research findings concerning the DAP increases the accuracy of their judgments; (d) that providing clinicians with an extended series of drawings enables them to in-

crease the accuracy of their judgments; and (e) that there is a positive correlation between a clinician's confidence in his judgments and the correctness of those judgments.

Major findings were: (a) that while art quality is the major factor influencing clinicians' adjustment ratings of DAP protocols, the actual patient-nonpatient status of the S is a factor which is also significantly related to evaluations of adjustment; (b) that artistically poor drawings appear to be judged more reliably and validly than those of better artistic quality; (c) that pathology itself does not result in poor art for a more personally meaningful type of drawing; (d) that use of DAP research findings does not result in improved accuracy of clinical judgments; (e) that clinical evaluations based on an extended series of drawings are significantly more accurate than those based on the usual two-drawing DAP; (f) that the benefits of an extended series seem due to its ability to offset the effects of artistic influence for those most influenced in their adjustment ratings by the art quality of the DAP; and (g) that while a clinician's confidence in his judgments relates significantly to the correctness of those judgments, it relates just as strongly to the assignment of patient status to a drawing and most strongly to levels of very good or very poor artistic quality.

It was felt that the results of the study, while reaffirming the strong influence of art quality on clinical judgment, did suggest that there are ways in which the effects of this variable might possibly be offset. It was suggested, however, that Ham-

mer's attempts to defend the two-drawing DAP could not adequately explain why clinicians' feelings of confidence in their evaluations were related much more strongly to the level of artistic quality than to the actual correctness of their judgments. The study noted also the importance of devising techniques to meet the needs of the individual clinician, since techniques which could be seen as helping one judge to improve on his evaluations would leave another judge relatively unaffected.

Abstract

41 patients (P) and 63 nonpatients (NP) took the DAP in an effort to study the relationship between artistic quality (AQ) and clinical judgment. An additional 20 P's and 20 NP's submitted an extended drawing series (EDS). Drawings were rated for AQ by 5 art judges while judged for adjustment and P-NP status by 5 "naïve" and 5 "informed" graduate clinicians who also rated confidence in their judgments. AQ mainly accounted for the variance of adjustment ratings ($p < .0001$) though true P-NP status was involved as well ($p < .05$). Confidence ratings were significantly related to correctness of judgment ($r_b = .24$), to the assignment of P status ($r_b = .22$), and most strongly to very good and very poor AQ ($\eta^2 = .56$). Use of EDS resulted in greater judgmental accuracy ($p < .0001$) which seemed due to its ability to offset the effects of AQ. Artistically poor drawings were judged more reliably and validly than those of better AQ. Use of DAP research findings failed to improve clinical judgment. Results are discussed in relation to Hammer's criticisms of DAP research and in relation to clinical training practices.

References

- Albee, G. W., & Hamlin, R. M. Judgment of adjustment from drawings: The applicability of rating scale methods. Journal of Clinical Psychology, 1950, 6, 363-365.
- Apfeldorf, M., & Smith, W. The representation of the body self in human figure drawings. Journal of Projective Techniques and Personality Assessment, 1966, 30, 283-289.
- Bieliauskas, V. J., & Bristow, R. B. The effect of formal art training upon the quantitative scores of the H-T-P drawings. Journal of Clinical Psychology, 1959, 15, 57-59.
- Brown, D. G., & Tolor, A. Human figure drawing as indicators of sexual identification and inversion. Perceptual and Motor Skills, 1957, 7, 199-211.
- Bruning, J. L., & Kintz, B. L. Computational handbook of statistics. Glenview, Illinois: Scott, Foresman, 1968.
- Buck, J. N. The H-T-P technique: A qualitative and quantitative scoring manual. Journal of Clinical Psychology, 1948, 4, 317-396.
- Butler, R. L., & Marcuse, F. L. Sex identification at different ages using the Draw-A-Person Test. Journal of Projective Techniques, 1959, 23, 299-302.
- Caligor, L. The determination of the individual's unconscious conception of his masculinity-femininity identification. Journal of Projective Techniques, 1951, 15, 494-509.
- Caligor, L. The detection of paranoid trends by the eight card redrawing test (8 CRT). Journal of Clinical Psychology, 1952, 8, 397-401.
- Caligor, L. Quantification of the eight card redrawing test (8 CRT). Journal of Clinical Psychology, 1953, 9, 356-361.
- Craddick, R. A. The self-image in the Draw-A-Person test and self-portrait drawings. Journal of Projective Techniques and Personality Assessment, 1963, 27, 288-291.
- Duncan, D. B. Multiple range and multiple F tests. Biometrics, 1955, 11, 1-42.
- Edwards, A. L. Experimental design in psychological research. New York: Holt, Rinehart, & Winston, 1960.
- Feldman, M. J., & Hunt, R. G. The relation of difficulty in drawing to ratings of adjustment based on human figure drawings. Journal of Consulting Psychology, 1958, 22, 217-219.

- Fisher, S. Body reactivity gradients and figure drawing variables. Journal of Consulting Psychology, 1959, 23, 54-59.
- Grams, A., & Rinder, L. Signs of homosexuality in human figure drawings. Journal of Consulting Psychology, 1958, 22, 394.
- Guinan, J. F., & Hurley, J. R. An investigation of the reliability of human figure drawings. Journal of Projective Techniques, 1965, 29, 300-304.
- Guilford, J. Psychometric methods. New York: McGraw-Hill, 1954.
- Hammer, E. F. The clinical application of projective drawings. Springfield, Illinois: Thomas, 1958.
- Hammer, E. F. Critique of Swenson's "Empirical evaluations of human figure drawing." Journal of Projective Techniques, 1959, 23, 30-31.
- Hammer, E. F. Projective drawings. In A. Rabin (Ed.) Projective techniques in personality assessment. New York: Springer, 1968.
- Hammer, E. F. DAP: Back against the wall? Journal of Consulting and Clinical Psychology, 1969, 33, 151-156.
- Hammer, M., & Kaplan, A. M. The reliability of sex of first drawn figure by children. Journal of Clinical Psychology, 1964, 20, 251-252.
- Hammer, M., & Kaplan, A. M. The reliability of childrens' human figure drawings. Journal of Clinical Psychology, 1966, 22, 316-319.
- Handler, L., & Reyher, J. The effects of stress on the Draw-A-Person test. Journal of Consulting Psychology, 1964, 28, 259-264.
- Handler, L., & Reyher, J. Figure drawing anxiety indices: A review of the literature. Journal of Projective Techniques, 1965, 29, 305-313.
- Handler, L., & Reyher, J. Relationship between GSR and anxiety indexes in projective drawings. Journal of Consulting Psychology, 1966, 30, 60-67.
- Harris, D. B. Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace, & World, 1963.
- Harrower, M. R. The most Unpleasant Concept Test: A graphic projective technique. Journal of Clinical Psychology, 1950, 6, 213-233.

- Hiler, E. W., & Nesvig, D. An evaluation of criteria used by clinicians to infer pathology from figure drawing. Journal of Consulting Psychology, 1965, 29, 520-529.
- Hunt, R. G., & Feldman, M. J. Body image and ratings of adjustment on human figure drawings. Journal of Clinical Psychology, 1960, 16, 35-38.
- Jones, L. W., & Thomas, C. P. Studies on figure drawings: A review of the literature (1949-1959). Psychometric Quarterly Supplement, 1961, 35, 212-261.
- Kamano, D. K. An investigation of the meaning of human figure drawing. Journal of Clinical Psychology, 1960, 16, 429-430.
- Katkov, B., & Goodman, M. The Draw-a-Person tests of obese women. Journal of Clinical Psychology, 1953, 9, 362-364.
- Kinget, G. M. The Drawing Completion Test--a projective technique for the investigation of personality. New York: Grune & Stratton, 1952.
- Lair, C. V., & Trapp, E. P. Performance decrement of the H-T-P Test as a function of adjustment level. Journal of Clinical Psychology, 1960, 16, 431.
- Levy, B., Lomax, J., & Minsky, R. An underlying variable in the clinical evaluation of drawings of human figures. Journal of Consulting Psychology, 1963, 27, 508-512.
- Lewinsohn, P. M. Psychological correlates of overall quality of figure drawings. Journal of Consulting Psychology, 1965, 29, 504-512.
- Litt, S., & Margolies, A. Sex change in successive Draw-A-Person tests. Journal of Clinical Psychology, 1966, 22, 471.
- Machover, K. Personality projection in the drawing of the human figure. Springfield, Illinois: Thomas, 1949.
- Marais, H. C., & Strumpfer, D. J. DAP body image disturbance scale and quality of drawing. Perceptual and Motor Skills, 1965, 21, 196.
- McNemar, Q. Psychological statistics. New York: Wiley, 1962.
- Milikian, L. H., & Wahab, A. Z. First-drawn picture: A cross-cultural investigation. Journal of Projective Techniques and Personality Assessment, 1969, 33, 539-541.

- Murray, D. C., & Deabler, H. L. Drawings, diagnosis, and the clinicians learning curve. Journal of Projective Techniques, 1958, 22, 415-420.
- Murstein, B. Handbook of projective techniques. New York: Basic Books, 1965.
- Nichols, R. C., & Strumpfer, D. J. Factor analysis of draw-a-person test scores. Journal of Consulting Psychology, 1962, 26, 156-161.
- Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1967.
- Prater, G. F. A comparison of the head and body size in the drawing of the human figure by hemiplegic and nonhemiplegic persons. Unpublished masters thesis, University of Kentucky, 1950.
- Roback, H. B. Human figure drawings: Their utility in the clinical psychologist's armamentarium for personality assessment. Psychological Bulletin, 1968, 70, 1-19.
- Schaeffer, R. W. Clinical psychologists' ability to use the D-A-P Test as an indicator of personality adjustment. Journal of Consulting Psychology, 1964, 28, 383.
- Schmidt, L., & McGowan, J. The differentiation of human figure drawings. Journal of Consulting Psychology, 1959, 23, 129-133.
- Sherman, L. J. The influence of artistic quality on judgments of patient and non-patient status from human figure drawings. Journal of Projective Techniques, 1958, 22, 338-340. (a)
- Sherman, L. J. Sexual differentiation or artistic ability? Journal of Clinical Psychology, 1958, 14, 170-171. (b)
- Silverstein, A. B., & Robinson, H. A. The representation of physique in children's figure drawings. Journal of Consulting Psychology, 1961, 25, 146-148.
- Simms, N. An analysis of the human figure drawings of orthopedic and non-orthopedic children. Unpublished masters thesis, University of Nebraska, 1951.
- Starr, S., & Marcuse, F. L. Reliability in the Draw-A-Person test. Journal of Projective Techniques, 1959, 23, 83-86.
- Stricker, G. Actuarial, naive, clinical, and sophisticated clinical prediction of pathology from figure drawings. Journal of Consulting Psychology, 1967, 31, 492-494.

- Sundberg, N. The practice of psychological testing in clinical services in the United States. American Psychologist, 1961, 16, 79-83.
- Swenson, C. H. Empirical evaluation of human figure drawings. Psychological Bulletin, 1957, 54, 431-466.
- Swenson, C. H. Empirical evaluations of human figure drawings: 1957-1966. Psychological Bulletin, 1968, 70, 20-44.
- Swenson, C. H., & Newton, K. The development of sexual differentiation on the DAP test. Journal of Clinical Psychology, 1955, 11, 417-419.
- Vilhouette, A. An investigation of the use of the DAP in the diagnosis of homosexuality in mentally deficient males. American Journal of Mental Deficiency, 1958, 62, 708-711.
- Wanderer, Z. W. Validity of clinical judgments based on human figure drawings. Journal of Consulting and Clinical Psychology, 1969, 33, 143-150.
- Watson, C. Relationship of distortion to DAP diagnostic accuracy among psychologists at 3 levels of sophistication. Journal of Consulting Psychology, 1967, 31, 142-146.
- White, W. An investigation of the relationship between difficulty of drawing and rated disturbance based on human figure drawings. Unpublished masters thesis, Loyola University, Chicago, 1969.
- Whitmyre, J. W. The significance of artistic excellence in the judgment of adjustment inferred from human figure drawings. Journal of Consulting Psychology, 1953, 17, 421-424.
- Woods, W., & Cook, W. Proficiency in drawing and placement of hands in drawings of the human figure. Journal of Consulting Psychology, 1954, 18, 119-121.
- Wysocki, B. A., & Whitney, E. Body image of crippled children as seen in draw-a-person test behavior. Perceptual and Motor Skills, 1965, 21, 499-504.

Approval Sheet

The dissertation submitted by James C. Young has been read and approved by members of the Department of Psychology.

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

September 16, 1970
Date

Frank K. Obler
Signature of Adviser